

University of Groningen

Nucleotide sequence of the *Agrobacterium tumefaciens* octopine Ti plasmid-encoded tmr gene

Heidekamp, F.; Dirkse, W.G.; Hille, J.; Ormondt, H. van

Published in:
Nucleic Acids Research

DOI:
[10.1093/nar/11.18.6211](https://doi.org/10.1093/nar/11.18.6211)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1983

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Heidekamp, F., Dirkse, W. G., Hille, J., & Ormondt, H. V. (1983). Nucleotide sequence of the *Agrobacterium tumefaciens* octopine Ti plasmid-encoded tmr gene. *Nucleic Acids Research*, 11(18). <https://doi.org/10.1093/nar/11.18.6211>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Nucleotide sequence of the *Agrobacterium tumefaciens* octopine Ti plasmid-encoded *tmr* gene

F.Heidekamp*, W.G.Dirkse, J.Hille¹ and H.van Ormondt²

Research Institute ITAL, P.O. Box 48, 6700 AA Wageningen, ¹Dept. Plant Molecular Biology, State University of Leiden, Biochemistry Building, Wassenaarseweg 64 and ²Dept. Medical Biochemistry, State University of Leiden, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands

Received 27 July 1983; Revised and Accepted 1 September 1983

ABSTRACT

The nucleotide sequence of the *tmr* gene, encoded by the octopine Ti plasmid from *Agrobacterium tumefaciens* (pTiAch5), was determined. The T-DNA, which encompasses this gene, is involved in tumor formation and maintenance, and probably mediates the cytokinin-independent growth of transformed plant cells. The nucleotide sequence of the *tmr* gene displays a continuous open reading frame specifying a polypeptide chain of 240 amino acids. The 5'- terminus of the polyadenylated *tmr* mRNA isolated from octopine tobacco tumor cell lines was determined by nuclease S1 mapping. The nucleotide sequence 5'-TATAAAA-3', which sequence is identical to the canonical "TATA" box, was found 29 nucleotides upstream from the major initiation site for RNA synthesis. Two potential polyadenylation signals 5'-AATAAA-3' were found at 207 and 275 nucleotides downstream from the TAG stopcodon of the *tmr* gene. A comparison was made of nucleotide stretches, involved in transcription control of T-DNA genes.

INTRODUCTION

Tumor-inducing (Ti) plasmids harbored by *Agrobacterium tumefaciens* cause neoplastic cell growth called crown galls on most dicotyledonous plants (for recent reviews see refs. 1,2). Tumor formation originates from the transfer of a specific Ti plasmid DNA segment, designated T-region, from the bacteria to the plant cells. Beside this T-region, a second portion of the Ti plasmid, the virulence region, is required for tumor induction (3,4). Upon transfer, the T-region is covalently integrated into plant nuclear DNA (5,6,7). Ti plasmid transformed plant cells are characterized by unlimited cell proliferation and the ability to grow in the absence of phytohormones like a cytokinin and an auxine. The tumor cells synthesize opines which are catabolized by the bacteria (8,9,10). Depending on the type of opine(s), produced in crown galls, the Ti plasmids have been classified in three major groups: octopine Ti plasmids, nopaline Ti plasmids and agropine Ti plasmids (11). In octopine tumors it has been shown that the enzyme octopine synthase (LpDH) is encoded by T-DNA

(12,13). These findings have prompted the development of Ti plasmid-derived vectors for the genetic manipulation of plant cells.

The extent of T-DNA and the T-DNA organization in octopine tumor tissues have been investigated (6,7,14). In octopine tumor cells the T-DNA consists of either one or two segments which originate from adjacent regions in the Ti plasmid DNA. All tumor cell lines investigated to this date contain the left part of the T-DNA (T_L -DNA), whereas the right part of the T-DNA (T_R -DNA) is present only in a limited number of tumor cell lines. Hybridization studies have shown that the T_L -DNA in octopine tumor cells encodes eight polyadenylated mRNAs (15,16,17). At least five and probably six of these transcripts have been shown to be involved in the process of tumor formation and maintenance (18,19,20,21).

In this paper the nucleotide sequence of the octopine T-region, encoding the tmr gene, is presented. This cistron specifies one of the above-mentioned transcripts (viz. transcript 4); presumably its product inhibits root formation of the tumors on certain plant species and appears to play a role in the cytokinin-independent growth of transformed cells (17,21). The 5'- terminus of the polyadenylated tmr mRNA, isolated from various octopine tumor cell lines, was determined by nuclease S1 mapping. A comparison is made between nucleotide stretches, involved in regulation of transcription of the tmr gene, of the octopine and nopaline synthase genes (22,24,25) and of the region which encodes octopine transcript 7 (23).

MATERIALS AND METHODS

Enzymes

The various restriction endonucleases indicated in this paper were obtained from New England Biolabs, Inc. (Beverly, MA) with the exception of BamHI, PstI, ClaI and HpaII which were purchased from Boehringer (Mannheim) and Sau3AI which was bought from Amersham International Ltd. (Amersham, UK). Assay conditions for restriction endonucleases were as described by the manufacturers. T_4 DNA ligase, T_4 polynucleotide kinase and calf intestine alkaline phosphatase were purchased from Boehringer (Mannheim); DNA polymerase I (large fragment) was from New England Biolabs, Inc. (Beverly, MA) and nuclease S1 was from Sigma (St. Louis, MO).

Recombinant plasmids and preparation of DNAs

Subfragments of the restriction DNA fragment EcoRI-7 from pTiAch5, on

Table 1: Characteristics of T-DNA containing recombinant plasmids, used for DNA sequence analysis.

Plasmid	Phenotype	Approximate size (bp)	Vector	T-DNA insert
pJH 189	Amp ^R , Tet ^S	5345	pBR 322	<u>Bam</u> HI-28 (28)
pJH 190	Amp ^S , Km ^R	3200	pACYC 177*	Left terminal <u>Bam</u> HI- <u>Pst</u> I subfragment of the restriction DNA fragment <u>Bam</u> HI-17 ^a (15,28)
pJH 191	Amp ^S , Tet ^R	5225	pBR 322	<u>Pst</u> I- <u>Pst</u> I subfragment of the restriction DNA fragment <u>Bam</u> HI-17 ^a (15,28)
pJH 589	Amp ^R , Tet ^R , Cm ^S	6250	pBR 329	<u>Hpa</u> I-14 (28,29)

* NB: pACYC 177 from which the BamHI-PstI restriction DNA fragment, which contains part of the Amp^R-gene, has been deleted.

which the tmr gene is located (7,20,26), were cloned according to standard procedures (27). The obtained recombinant plasmids, which were used for DNA sequence analysis, are listed in Table 1.

Plasmid DNA was isolated according to a modified cleared-lysate procedure (27). Supercoiled plasmid DNA was isolated by sucrose gradient centrifugation which was followed by CsCl buoyant-density centrifugation in the presence of 200 µg/ml ethidium bromide (30).

Purification and labelling of restriction DNA fragments, chemicals used and DNA sequence analysis

Recombinant plasmids were digested with appropriate restriction endonucleases and the digestion products were separated on horizontal 1% agarose gels in 50 mM Tris-borate (pH 8.3), 1 mM EDTA containing 1 µg/ml ethidium bromide. T-DNA inserts were excised from these gels and electroeluted at 4 °C in 20 mM Tris-acetate (pH 7.8), 10 mM Na acetate, 1 mM EDTA for 16 hr at 100 V. After mixing the eluate with n-butanol, the DNA fragments were precipitated with ethanol. Prior to labelling or secondary cleavage with restriction endonucleases, the restriction DNA fragments were purified by gel filtration through a 1-ml Sephadex G-50 medium column [Bio-Rad laboratories (Richmond,CA)] column in 10 mM Tris-HCl (pH 7.6), 1 mM EDTA.

[α-³²P] dNTPs and [γ-³²P] ATP were purchased from New England Nuclear (Boston, MA). End-labelling of DNA fragments was performed according to standard procedures (27). Nucleotide sequence analysis of 3'- or 5'-

labelled restriction fragments was performed according to the chemical degradation method (31,32). Nucleotide sequence data were processed using computer programs originally developed by Staden (33).

Nuclease S1 mapping

All RNA preparations, used in this study, were generous gifts of Dr. J.H.C. Hoge (Dept. of Plant Molecular Biology, State University of Leiden). The plant tissue culture lines (7) from which these RNAs were isolated were octopine tumor cell line B6S3 (which contains T-DNA fragment EcoRI-7) and 4013-2 (which contains EcoRI-7 and EcoRI-32). In control experiments total RNA isolated from the plasmid-cured Agrobacterium tumefaciens strain LBA 4011 was used.

Nuclease S1 mapping experiments were performed as described by Berk and Sharp (34) with the following modifications. ³²P-end-labelled T-DNA probes were mixed with polyadenylated tumor tissue RNA (20 µg) or bacterial RNA (20 µg) in 20 µl hybridization buffer containing 80% formamide, 40 mM PIPES (pH 6.4), 400 mM NaCl, 1 mM EDTA and subsequently incubated for 15 min at 90 °C. Hybridization was performed for 16 hr at 50 °C after which the incubation was stopped by quick chilling of the samples in a dry-ice/ethanol bath. The samples were diluted 10 times with nuclease S1 buffer containing 250 mM NaCl, 30 mM Na acetate, 1 mM ZnSO₄, 20 µg/ml sonicated calf thymus DNA and incubated with 200 units/ml nuclease S1 for 45 min at 20 °C. Nuclease S1-resistant hybrids were precipitated twice with ethanol and subsequently analysed on 5% or 8% polyacrylamide gels in 7 M urea next to ³²P-labelled restriction DNA fragments of known size or next to chemical degradation products of the ³²P-labelled T-DNA probe used.

RESULTS

The T_L-DNA in octopine crown gall tissue codes for eight polyadenylated transcripts (16,17). By site-directed mutagenesis of the T_L-region most of the T_L-DNA-derived transcripts could be ascribed to different loci (18,19,20,21). These studies revealed that the T_L-DNA part, which suppresses root formation of the tumors, encodes transcript 4 (approx. size 1200 nucleotides, poly(A) tail included). The transcript 4-encoding region encompasses part of the restriction DNA fragment BamHI-28 and part of the adjacent restriction DNA fragment BamHI-17^a of pTiAch5, assuming that no splicing occurs (3,16,28). Therefore, nucleotide sequence

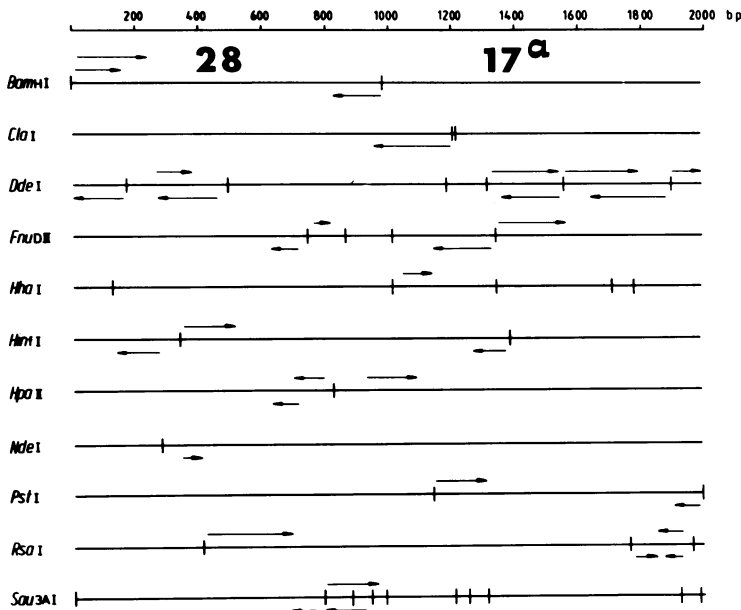


FIGURE 1: Diagram of the analysed restriction DNA fragments in the tmr gene region. Vertical bars represent restriction endonuclease cleavage sites; the length of the arrows indicate the length of the nucleotide stretches, deduced from the various restriction DNA fragments. The arrowheads indicate the direction of nucleotide sequence elucidation for the corresponding restriction DNA fragments. Note that physical maps of the restriction DNA fragment BamHI-28 and of part of the adjacent restriction DNA fragment BamHI-17^a are shown.

analysis of this region should reveal the nucleotide sequence of the tmr gene.

Nucleotide sequence analysis of the tmr gene region

In order to determine the nucleotide sequence of the tmr gene region, subfragments of the restriction fragment EcoRI-7 from pTiAch5 were cloned in Escherichia coli vectors as indicated in Materials and Methods. The T-DNA inserts present in the recombinant plasmids pJH 189, pJH 190 and pJH 191 (see Table 1), which originate from adjacent regions on pTiAch5, were isolated and the complete nucleotide sequence of these DNA fragments was elucidated according to a sequence strategy shown in Fig. 1.

In order to be certain that joining of the nucleotide sequences, deduced from the T-DNA inserts present in the recombinant plasmids pJH 189, pJH 190 and pJH 191, yielded the complete nucleotide sequence of the tmr gene

GGATCCTGTTACAAGTATTGCAAGTTTTATAAATTGCATATTAATGCAATCTTGATTTTTAAACAACGAACGTAAT
1
GGCGTAAAAATGTATGTTATATTATTTATATTTAATTATATTGGAGTGGCCATAATATGATGATTTTATAATTA
76
AAAAATATTTACTGTACATTGACTGAGATGGCACTGTTATTTCAACCATGAAATTTTGTGATTTTTTACAAT
151
AACATAATTGCAGGAAGTAAATAATAGACGCCGTGTTAAAAAATTGCAATCATATGTGCCTAACTATAGGGAC
226
AATTAAGTCAATGTGAATAGTCTCCCTTATTTTAACGACTCACCTAATCAAGTATTACAAAATATCTCACTTTTC
301
GTCAGTAATGATGTAATCAGAAGTGAATAGTACAAGTAAACCGTGAAAAACGTCATAGAGTGGCATGATTATAT
376
TCCTCTGCATTGCCAATTTATTACGCTTTATTGACTTAGGTGTGCCTTCGTTAGCGACAAATTGCTTTCAAGGA
451
GACAGCCATGCCCCACACTTTGTTGAAAACAAGTTGCCTTTTGGGATACGGTAAAGCCAGTTGCACCTTCAATAA
526
TGAATTTCAAGGAGACAAATATAACCGCCTCTGATAACACAATTCTCTAAATATAAAATCAGTTTGTATTCAATAT
601
MET ASP LEU HIS LEU ILE PHE GLY PRO THR CYS THR GLY LYS THR
ACTGCAAAAACTT ATG GAC CTG CAT CTA ATT TTC GGT CCA ACT TGC ACA GGA AAG ACG
676
THR THR ALA ILE ALA LEU ALA GLN GLN THR GLY LEU PRO VAL LEU SER LEU ASP ARG
ACG ACC GCG ATA GCT CTT GCC CAG CAG ACA GGG CTT CCA GTC CTT TCG CTT GAT CGG
735
VAL GLN CYS CYS PRO GLN LEU SER THR GLY SER GLY ARG PRO THR VAL GLU GLU LEU
GTC CAA TGC TGT CCT CAA CTA TCA ACC GGA AGC GGA CGA CCA ACA GTG GAA GAA CTG
792
LYS GLY THR THR ARG LEU TYR LEU ASP ASP ARG PRO LEU VAL GLU GLY ILE ILE ALA
AAA GGA ACG ACG CGT CTC TAC CTT GAT GAT CGG CCT CTG GTG GAG GGT ATC ATC GCA
849
ALA LYS GLN ALA HIS HIS ARG LEU ILE GLU GLU VAL TYR ASN HIS GLU ALA ASN GLY
GCC AAG CAA GCT CAT CAT AGG CTG ATC GAG GAG GTG TAT AAT CAT GAG GCC AAC GGC
906
GLY LEU ILE LEU GLU GLY GLY SER THR SER LEU LEU ASN CYS MET ALA ARG ASN SER
GGG CTT ATT CTT GAG GGA GGA TCC ACC TCG TTG CTC AAC TGC ATG GCG CGA AAC AGC
963
TYR TRP SER ALA ASP PHE ARG TRP HIS ILE ILE ARG HIS LYS LEU PRO ASP GLN GLU
TAT TGG AGT GCA GAT TTT CGT TGG CAT ATT ATT CGC CAC AAG TTA CCC GAC CAA GAG
1020
THR PHE MET LYS ALA ALA LYS ALA ARG VAL LYS GLN MET LEU HIS PRO ALA ALA GLY
ACC TTC ATG AAA GCG GCC AAG GCC AGA GTT AAG CAG ATG TTG CAC CCC GCT GCA GGC
1077
HIS SER ILE ILE GLN GLU LEU VAL TYR LEU TRP ASN GLU PRO ARG LEU ARG PRO ILE
CAT TCT ATT ATT CAA GAG TTG GTT TAT CTT TGG AAT GAA CCT CGG CTG AGG CCC ATT
1134
LEU LYS GLU ILE ASP GLY TYR ARG TYR ALA MET LEU PHE ALA SER GLN ASN GLN ILE
CTG AAA GAG ATC GAT GGA TAT CGA TAT GCC ATG TTG TTT GCT AGC CAG AAC CAG ATC
1191
THR ALA ASP MET LEU LEU GLN LEU ASP ALA ASN MET GLU GLY LYS LEU ILE ASN GLY
ACG GCA GAT ATG CTA TTG CAG CTT GAC GCA AAT ATG GAA GGT AAG TTG ATT AAT GGG
1248
ILE ALA GLN GLU TYR PHE ILE HIS ALA ARG GLN GLN GLU GLN LYS PHE PRO GLN VAL
ATC GCT CAG GAG TAT TTC ATC CAT GCG CGC CAA CAG GAA CAG AAA TTC CCC CAA GTT
1305
ASN ALA ALA ALA PHE ASP GLY PHE GLU GLY HIS PRO PHE GLY MET TYR ***
AAC GCA GCC GCT TTC GAC GGA TTC GAA GGT CAT CCG TTC GGA ATG TAT TAG GTTACGC
1362

CAGCCCTGCGTCGCACCTGTCTTCATCTGGATAAGATGTTGTAATTGTTTTGGCTTTGTCTGTGTGGCAGG
 1420
 GCGGCAAACTACTCCGACAATCCATCGTGTCTTCAAACCTTATGCTGGTGAACAAGTCTTAGTTTCCACGAAAGT
 1495
 ATTATGTTAAATTTTAAATTTTCGATGTATAATGTGGCTATAATTGTAAAAATAAACTATCGTAAGTGTGCGTGT
 1570
 TATGTATAATTTGTCTAAATGTTTAAATATATATCATAGAACGCAATAAAATATTAATATAGCGCTTTTATGAAAT
 1645
 ATAAATACATCATTACAAGTTGTTTATATTTTCGGGTACCTTTTCCATTATTTTTCGCAACAAGTCACGGATATTC
 1720
 GTGAAAACGACAAAACTGCGAAATTTGCGGGCAGTGCCTTCAGTTTTCCTATTAATATTTAGTTTGACACCACT
 1795
 TGCTATCATTCGCGCCAAGCTCAGCTGTTTCTTTTCTTGAAACGATGGATCGAATGAGCATGGCTCGGCAAGGTT
 1870
 GGCTGTACCATGTCTTCTCATGGCAAAGATGATCAACTGCAG
 1945

FIGURE 2: Nucleotide sequence of the tmr gene region. The noncoding (or sense) strand is displayed. The amino acid sequence of the tmr protein, deduced from the nucleotide sequence of the tmr gene, is indicated. Nucleotide stretches, which presumably are involved in initiation of transcription, are boxed. The G-residue at position 679 marks the major transcription initiation site. Putative polyadenylation signals are underlined. See text for further explanation.

region the following experiments were performed. The pTiAch5 restriction DNA fragment HpaI-14, which overlaps the internal BamHI-PstI fragment indicated in Fig. 1 (15,28) was isolated from the recombinant plasmid pJH 589 (see Table 1). Subsequently, this fragment was redigested with HpaII and ClaI, which was followed by 5'-end-labelling of the restriction DNA subfragments. The HpaII-ClaI restriction DNA fragment with an approximate length of 385 bp, which contains the internal BamHI-PstI fragment indicated in Fig. 1, was isolated and subjected to strand separation. Nucleotide sequence analysis of the separate DNA strands revealed the nucleotide sequences at the BamHI and PstI restriction sites, which sites had been used to generate the respective T-DNA inserts of the recombinant plasmids pJH 189, pJH 190 and pJH 191 (results not shown). The deduced nucleotide sequence of the tmr gene region, together with the amino acid sequence encoded by the tmr gene, is shown in Fig. 2.

Nuclease S1 mapping of the 5'-terminus of the tmr mRNA

In order to identify nucleotide sequences, involved in the regulation of transcription of the tmr gene region, nuclease S1-protection studies were performed. The 5'-terminus of the tmr mRNA was determined as follows. The restriction fragment BamHI-28 (N 1-981 in Fig. 2) was 5'-end-labelled with polynucleotide kinase and [γ -³²P] ATP and subsequently digested with

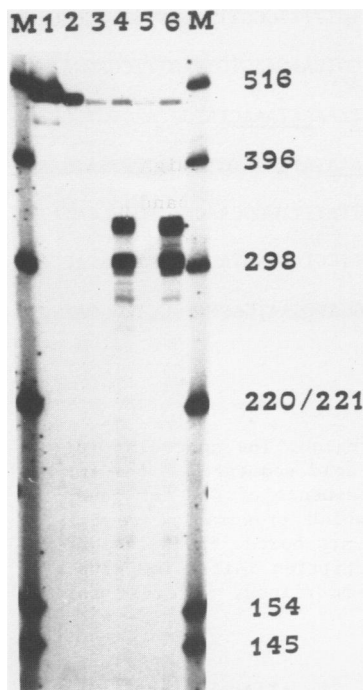


FIGURE 3:

Nuclease S1 mapping of the 5'-terminus of the octopine *tmr* mRNA. The autoradiograph of the denaturing gel shows: lane 1: DdeI-BamHI probe (N 487-981 in Fig. 2), without further treatment; lane 2: DdeI-BamHI probe, subjected to hybridization and nuclease S1 digestion in the absence of RNA; lanes 3 and 5: products from the hybridization of total RNA from *Agrobacterium tumefaciens* strain LBA 4011 against the DdeI-BamHI probe, obtained after nuclease S1 digestion (controls); lanes 4 and 6: products from the hybridization of octopine 4013-2 poly(A)⁺ RNA (lane 4) and B6S3 poly(A)⁺ RNA (lane 6) against the DdeI-BamHI probe, obtained after nuclease S1 digestion. In the lanes marked M, the ³²P-labelled HinfI fragments of pAT153 were separated (lengths indicated in nucleotides).

DdeI. The DdeI-BamHI subfragment (N 487-981) was isolated and hybridized against octopine tumor poly(A)⁺ RNA. After nuclease S1 digestion, the products were separated in a 5% polyacrylamide -7M urea gel next to ³²P-labelled restriction DNA fragments of known size. The results are shown in Fig. 3.

As shown in Fig. 3 (lanes 4 and 6), two fragments with an approximate length of 300 nucleotides and 335 nucleotides were protected from nuclease S1 digestion, which suggests that initiation of transcription of the *tmr* gene occurs at two different regions. From the length of these fragments it was deduced that the 5'-terminus of the *tmr* mRNA is located in the nucleotide stretch 5'-ACTGCAAA-3' (N 676-683 in Fig. 2) and in the nucleotide stretch 5'-TTCTCTAA-3' (N 642-649).

In order to define the 5'-terminus of the *tmr* mRNA more accurately, nuclease S1 mapping experiments were performed using the DdeI-HpaII restriction DNA fragment (N 487-817 in Fig. 2), 5'-³²P-labelled at the HpaII-end, as DNA probe. The products from the hybridization of octopine poly(A)⁺ RNA against the DdeI-HpaII probe were separated after nuclease S1 digestion in an 8% polyacrylamide-7 M urea gel next to the chemical

degradation products of this DNA fragment. Autoradiography of this gel revealed the presence of two clusters of nuclease S1-protected radioactive bands which comigrated with individual residues of the nucleotide stretch 5'-GCAAA-3' (N 679-683 in Fig. 2) and of the nucleotide stretch 5'-CTAATA-3' (N 646-651; results not shown). The relatively high amount of radioactivity present in the nuclease S1-protected band which comigrated with the band corresponding to the G-residue at position 679 suggests that initiation of transcription of the tmr mRNA preferably occurs at this site. Inspection of the nucleotide sequence of the tmr gene region upstream from this major mRNA initiation site (the "+1"-site) revealed the nucleotide stretch 5'-TATAAAA-3' at -29 to -23 nucleotides. This nucleotide sequence is identical to the "TATA" box sequence found in most eukaryotic genes (36) and probably functions as part of the promoter for initiation of transcription by the plant nuclear RNA polymerase II.

DISCUSSION

The Agrobacterium tumefaciens Ti plasmids are attractive vectors for the genetic manipulation of higher plants because of their natural ability to transform plant cells by inserting the T-region into chromosomal DNA of the host plant. However, transfer of the T-region to plant cells is accompanied by oncogenic transformation of these cells, which prevents their regeneration to intact, healthy plants. In order to construct Ti plasmid-derived vectors, which lack the T-DNA parts involved in tumorous growth of transformed plant cells, the extent of these regions has to be determined at the nucleotide level. In addition, nucleotide sequence analysis of these regions might reveal how expression of the plasmid-borne T-DNA genes is regulated in plant cells. In this paper the nucleotide sequence of the Agrobacterium tumefaciens pTiAch5 tmr gene region is presented (see Fig. 2). The nucleotide sequence of the tmr gene which is flanked by stretches rich in A-T bases, specifies a polypeptide chain of 240 amino acids with a predicted molecular weight of 27,003 Daltons. Presumably, the protein product of the tmr gene mediates the cytokinin-independent growth of transformed plant cells. This would suggest that the tmr protein is an enzyme, involved in the biosynthesis of cytokinins in these cells.

The region, shown in Fig. 2, is transcribed in plant cells (16,17) but the protein product has not yet been identified at this level. Expression of a 27,000 Daltons protein from this region has been observed in

Escherichia coli minicells and in cell-free systems, prepared from Agrobacterium tumefaciens and Escherichia coli (35). It is not clear whether the tmr protein and this 27,000 Daltons protein are related in structure and function.

The 5'-terminus of the tmr mRNA was determined by nuclease S1 mapping. The results suggest that initiation of transcription of the tmr gene occurs at multiple sites in the nucleotide stretch 5'-GCAAA-3' (N 679-683 in Fig. 2) and in the nucleotide stretch 5'-CTAATA-3' (N 646-651). The major RNA start site is presumably located at the G-residue which is found 11 nucleotides upstream from the ATG initiation codon for protein synthesis (N 679 in Fig. 2). Inspection of the nucleotide sequence, upstream from the major RNA start site revealed the nucleotide stretch 5'-TATAAAA-3' at -29 to -23 nucleotides, whereas at a similar distance upstream from the minor RNA initiation region the nucleotide stretch 5'-AATATAA-3' (N 617-623) is found.

Both nucleotide stretches show (almost) perfect homology to the "TATA" box sequence found in most eukaryotic genes which sequence functions as part of the promoter for initiation of transcription. It is attractive to investigate whether site-directed mutagenesis or deletion of either one of these possible promoter sequences affects the efficiency of transcription of the tmr gene.

Although the "TATA" box is necessary and sufficient for accurate initiation of transcription in vitro, regions further upstream have been proposed to be required for efficient in vivo transcription (37). One of these regions, the "CAAT" box (CONSENSUS: 5'-GG^C_TCAATCT-3'), is located at 70 to 80 nucleotides upstream from the RNA start site. Therefore, the region upstream the 5'-side of the tmr gene, together with the corresponding regions in the recently published nucleotide sequences of octopine synthase (ocs; 22), of nopaline synthase (nos; 24,25) and of the octopine transcript 7 gene (ot7; 23) have been inspected for nucleotide stretches which bear homology to the "CAAT" box (shown in Fig. 4).

As shown in Fig. 4, there is substantial sequence homology with the canonical "CAAT" box in the -80/-70 upstream region of the nopaline synthase gene, whereas sequence homology with the "CAAT" box in the corresponding regions of the octopine tmr gene, of the octopine synthase gene and of the gene, encoding octopine transcript 7, is poor. This would suggest that the -80/-70 region at the 5'-side of octopine T-DNA genes is not required for in vivo transcription. It should be noticed, however,

-80/-70 -34/-26

```

CONSENSUS:   GGCTCAATCT           TATAATAT
tmr          :   aaTgAATtT - 42 - TATAAAA - 22 - G - 10 - ATG
               -80                               1
               aGaCAATaT
               -66
               tGCCAATtT
               -218
ocs          :   aGTtAAagg - 35 - TATtTAA - 25 - A - 25 - ATG
               -76                               1
               aGTCAATac
               -58
               aGgCAATtT
               -160
nos          :   GGTCAcTaT - 43 - cATAAAT - 20 - A - 34 - ATG
               -79                               1
               cGaCAATCT
               -274
               cGCCAATaT
               -323
oT7          :   aaTtAAgCc - 40 - TATATAg - 23 - A - 15 - ATG
               -79                               1
               GtTCAAgCT
               -92

```

FIGURE 4: Nucleotide sequences upstream the 5' side of the tmr gene, of the octopine synthase gene, of the nopaline synthase gene and of the gene encoding octopine transcript 7 which may be involved in regulation of transcription. The position of the respective (major) RNA start sites was set to +1. Nucleotides which are homologous with part of the "CAAT" and "TATA" boxes are represented by upper-case characters; non-homologous nucleotides are represented by lower-case characters.

that insertion of Transposon 5 at a position, located 120 nucleotides upstream of the octopine synthase mRNA start site, abolishes octopine synthesis in tumor cells (22). As indicated in Fig. 4, nucleotide stretches upstream of the respective octopine T-DNA genes at other positions than the -80/-70 region bear more homology to the "CAAT" box sequence, whereas no other conserved nucleotide sequences are present in the regions upstream from the "TATA" box of octopine T-DNA genes. Therefore, it cannot be excluded that other "CAAT" stretches are involved in regulation of transcription of octopine T-DNA genes.

As shown in Fig. 2, two potential polyadenylation signals 5'-AATAAA-3' are present in the 3'-flanking region of the tmr gene at respectively 207

and 275 nucleotides downstream of the TAG stopcodon. Both AATAAA-stretches are preceded by a potential hairpin sequence (N 1578-1591 and N 1666-1681 in Fig. 2) which structures might be involved in the regulation of transcription-termination.

In conclusion it is clear that transcription of the bacterial T-DNA genes in plant cells is regulated by control sequences of eukaryotic nature. Further investigations on the identification of transcriptional control sequences in the T-DNA will be necessary to make the Ti-derived plasmids more suitable as vectors for the genetic engineering of plants.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. J.H.C. Hoge for gifts of RNA preparations and Mr. E.T.A. Zwemmer for his assistance in computer processing of nucleotide sequence data.

The work described in this paper is an initiation of cooperative research of the Dept. of Plant Molecular Biology, State University of Leiden (Head: Prof. R.A. Schilperoort) and the Research Institute ITAL (Prof. B. de Groot).

*To whom correspondence should be addressed

ABBREVIATIONS

Amp,	ampicillin
Tet,	tetracyclin
Km,	kanamycin
R or S,	resistance or sensitivity
Cm,	chloramphenicol
N,	nucleotide
bp,	base pair

REFERENCES

1. Bevan, M.W. and Chilton, M.-D. (1982) *Ann. Rev. Genet.* **16**, 357-384.
2. Relevant chapters in: "Molecular Biology of Plant Tumors", Kahl, G. and Schell, J.S. eds. (1982), Academic Press, New York.
3. Ooms, G., Klapwijk, P.M., Poulis, J.A. and Schilperoort, R.A. (1980) *J. Bact.* **144**, 82-91.
4. Hille, J., Klasen, I. and Schilperoort, R. (1982) *Plasmid* **7**, 107-118.
5. Chilton, M.-D., Drummond, M.H., Merlo, D.J., Sciaky, D., Montoya, A.L., Gordon, M.P. and Nester, E.W. (1977) *Cell* **11**, 263-271.
6. Thomashow, M.F., Nutter, R., Montoya, A.L., Gordon, M.P. and Nester, E.W. (1980) *Cell* **19**, 729-739.
7. Ooms, G., Bakker, A., Molendijk, L., Wullems, G.J., Gordon, M.P., Nester, E.W. and Schilperoort, R.A. (1982) *Cell* **30**, 589-597.

8. Bomhoff, G., Klapwijk, P.M., Kester, H.C.M., Schilperoort, R.A., Hernalsteens, J.P. and Schell, J. (1976) *Mol. Gen. Genet.* 145, 177-181.
9. Montoya, A.L., Chilton, M.-D., Gordon, M.P., Sciaky, D. and Nester, E.W. (1977) *J. Bact.* 129, 101-107.
10. Van Larebeke, N., Engler, G., Holsters, M., Van den Elsacker, S., Zaenen, I., Schilperoort, R.A. and Schell, J. (1974) *Nature* 252, 169-170.
11. Guyon, P., Chilton, M.-D., Petit, A. and Tempé, J. (1980) *Proc. Natl. Acad. Sci. USA* 77, 2693-2697.
12. Schröder, J., Schröder, G., Huisman, H., Schilperoort, R.A. and Schell, J. (1981) *FEBS Lett.* 129, 166-168.
13. Murai, N. and Kemp, J.D. (1982) *Proc. Natl. Acad. Sci. USA* 79, 86-90.
14. De Beuckeleer, M., Lemmers, M., De Vos, G., Willmitzer, L., Van Montagu, M. and Schell, J. (1981) *Mol. Gen. Genet.* 183, 283-288.
15. Gelvin, S.B., Thomashow, M.F., McPherson, J.C., Gordon, M.P. and Nester, E.W. (1982) *Proc. Natl. Acad. Sci. USA* 79, 76-80.
16. Willmitzer, L., Simons, G. and Schell, J. (1982) *The EMBO Journal* 1, 139-146.
17. Willmitzer, L., Dhaese, P., Schreier, P.H., Schmalenbach, W., Van Montagu, M. and Schell, J. (1983) *Cell* 32, 1045-1056.
18. Garfinkel, D.J., Simpson, R.B., Ream, L.W., White, F.F., Gordon, M.P. and Nester, E.W. (1981) *Cell* 27, 143-153.
19. Ream, L.W., Gordon, M.P. and Nester, E.W. (1983) *Proc. Natl. Acad. Sci. USA* 80, 1660-1664.
20. Leemans, J., Deblaere, R., Willmitzer, L., De Greve, H., Hernalsteens, J.P., Van Montagu, M. and Schell, J. (1982) *The EMBO Journal* 1, 147-152.
21. Joos, H., Inzé, D., Caplan, A., Sormann, M., Van Montagu, M. and Schell, J. (1983) *Cell* 32, 1057-1067.
22. De Greve, H., Dhaese, P., Seurinck, J., Lemmers, M., Van Montagu, M. and Schell, J. (1983) *J. Mol. Applied Genet.* 1, 499-511.
23. Dhaese, P., De Greve, H., Gielen, J., Seurinck, J., Van Montagu, M. and Schell, J. (1983) *The EMBO Journal* 2, 419-426.
24. Depicker, A., Stachel, S., Dhaese, P., Zambryski, P. and Goodman, H.M. (1982) *J. Mol. Applied Genet.* 1, 561-573.
25. Bevan, M., Barnes, W.M. and Chilton, M.-D. (1983) *Nucl. Acids Res.* 11, 369-385.
26. Ooms, G., Hooykaas, P.J.J., Molenaar, G. and Schilperoort, R.A. (1981) *Gene* 14, 33-50.
27. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in: "Molecular Cloning, a laboratory manual", CSHL, Cold Spring Harbor, New York.
28. De Vos, G., De Beuckeleer, M., Van Montagu, M. and Schell, J. (1981) *Plasmid* 6, 249-253.
29. Covarrubias, L. and Bolivar, F. (1982) *Gene* 17, 79-89.
30. Radloff, R., Bauer, W. and Vinograd, J. (1967) *Proc. Natl. Acad. Sci. USA* 57, 1514-1521.
31. Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 560-564.
32. Maxam, A.M. and Gilbert, W. (1980) in: *Methods in Enzymology*, Colowick, S.P. and Kaplan, N.O. eds. 69, 499-559; Academic Press, New York.
33. Staden, R. (1977) *Nucl. Acids Res.* 4, 4037-4051.
34. Berk, A.J. and Sharp, P.A. (1977) *Cell* 12, 721-732.
35. Schröder, G., Klipp, W., Hillebrand, A., Ehring, R., Koncz, C. and Schröder, J. (1983) *The EMBO Journal* 2, 403-409.
36. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* 50, 349-383.
37. Shenk, T. (1981) in: *Curr. Top. Microbiol. Immunol.* 93, 25-46.